**Predicting Weather Changes using ARIMA and SARIMA Time Series Model**

**Table of Contents:**

1. **Requirements**
2. **Dataset**
3. **Data Preprocess**
4. **Data Cleaning**
5. **Exploring and Visualization:**
6. **Approach One: ARIMA Model**
7. **Approach Two: SARIMA Model**
8. **Validation for both Models via Mean Squared Error (MSE)**
9. **Final Results for both Models**
10. **Requirements:**

· Python 3.5 or higher

· Numpy

· Pandas

· Matplotlib

· Statsmodels

· Adfuller

· Itertools

· Warnings

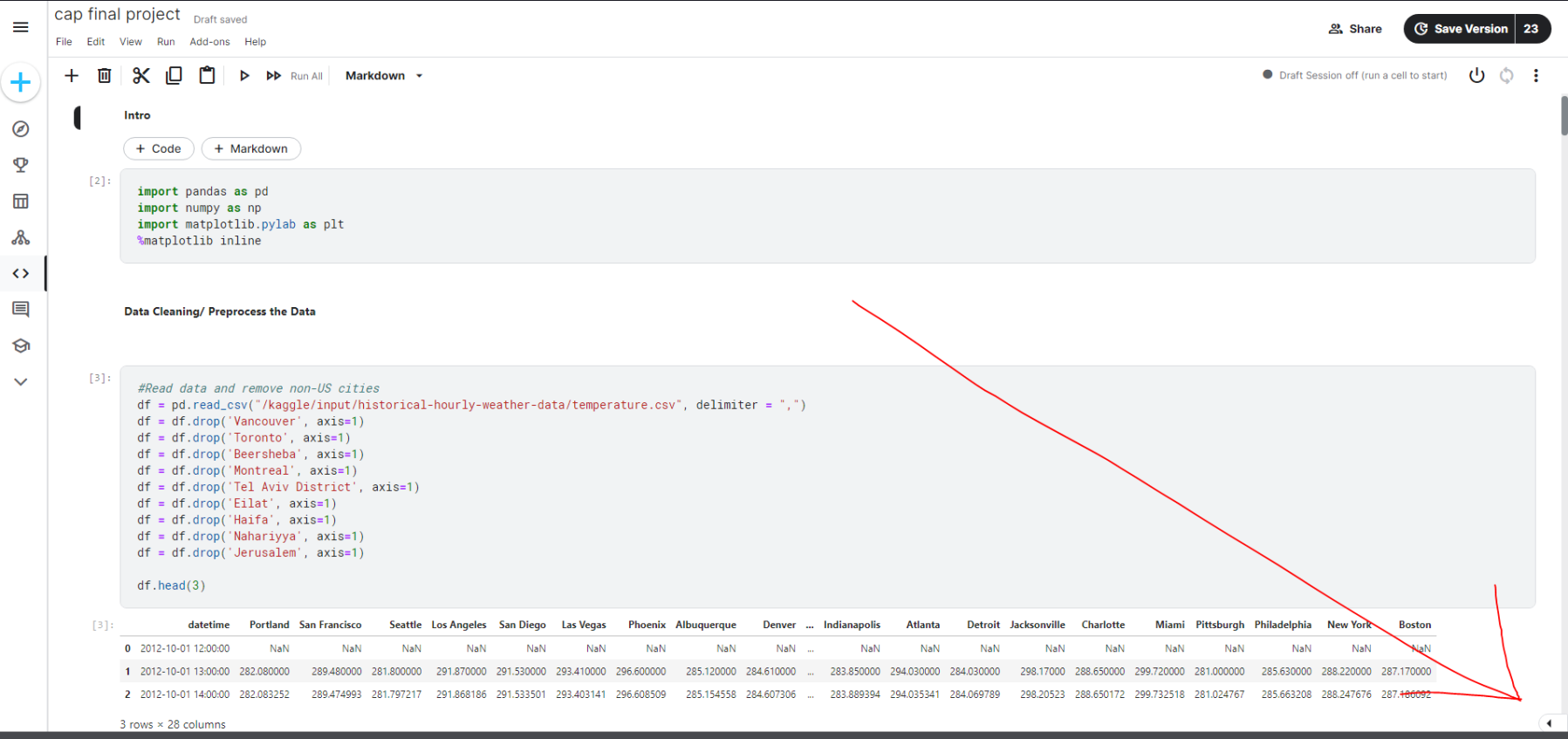
· datetime

· Pmdarima

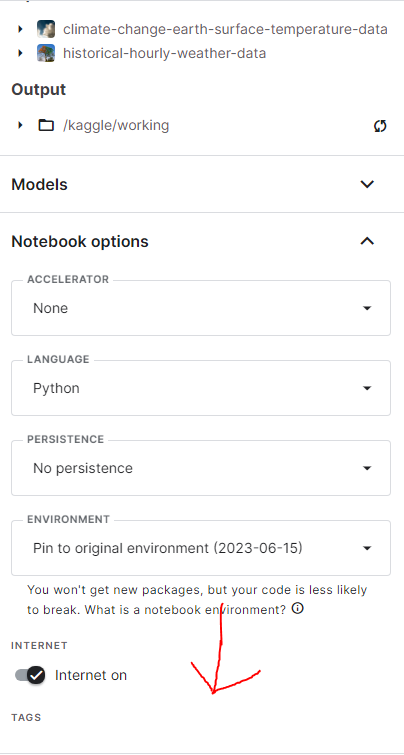
· Auto\_arima

**In order to enable Pmdarima and for the code to run correctly, internet must be enabled in your Kaggle notebook:**

Select the arrow on the bottom right hand corner of the notebook.

****

Toggle internet settings to on.

****

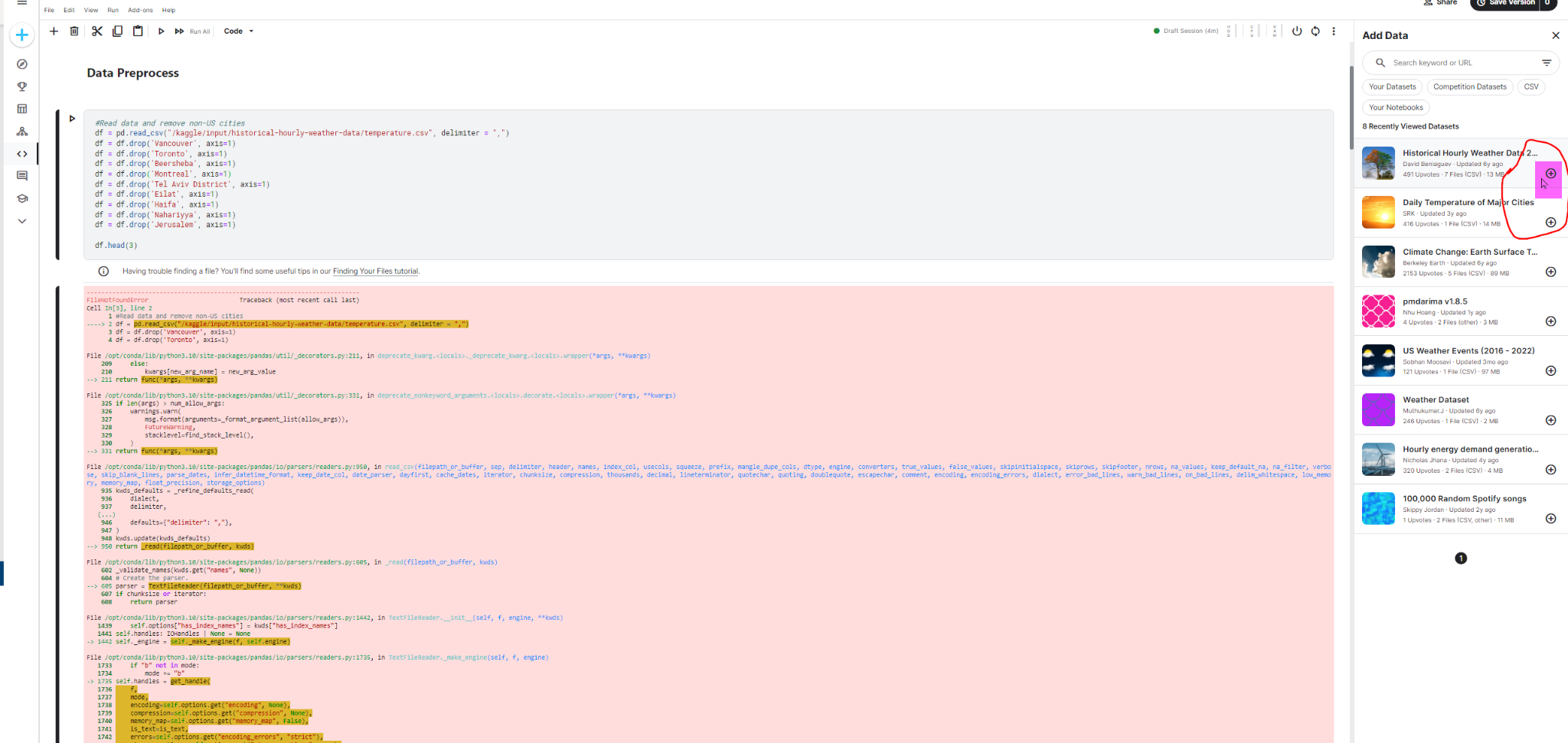
If the option to enable internet access is not available, there will be a link in its place allowing the user to verify their account with a phone number. The option should then be available.

**The data set will have to be added in order for the code to run.**

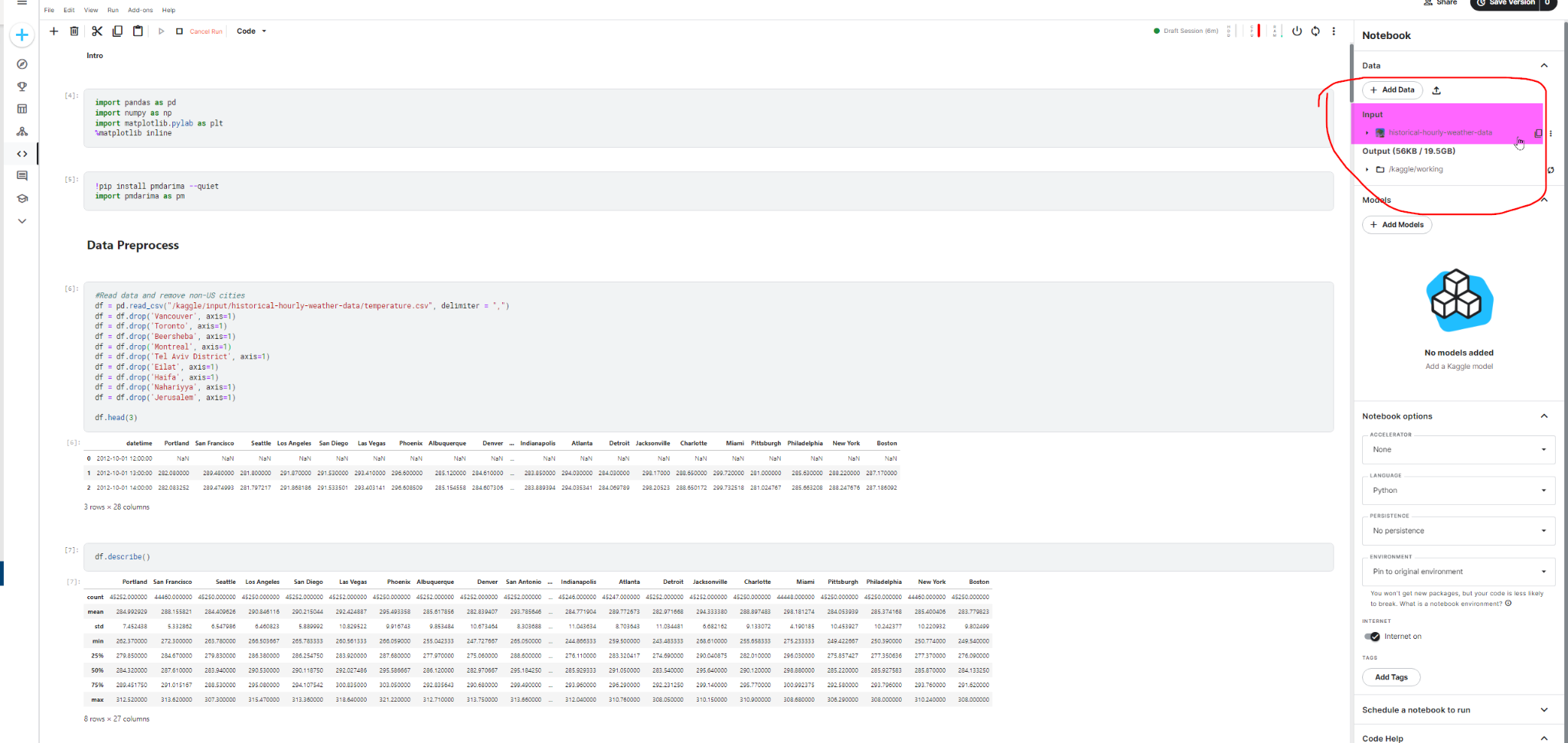
Click on ‘Add data’ in the right hand pane.



Click on the ‘+’ button next to the ‘Historic Hourly Weather Data’ dataset.



You should now see the dataset added under ‘Input’.

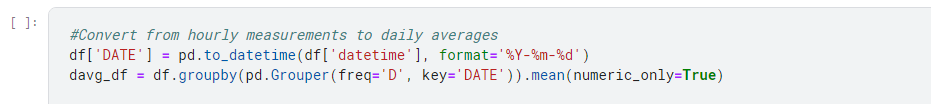
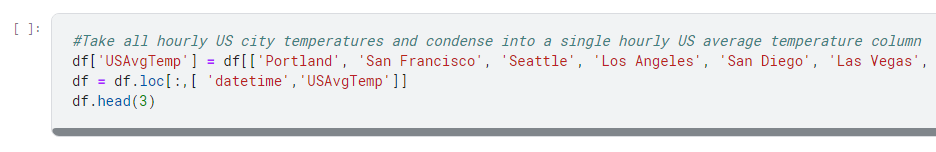


1. **Dataset:**

<https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data?select=temperature.csv>

1. **Data Preprocess:**

Pre-processing began with excluding all non-US cities from the dataframe and combining the rest into a single column for average hourly temperature in the US by finding the mean. This was followed by formatting the dates to remove the timestamp and then grouping all of the hourly measurements into daily averages.

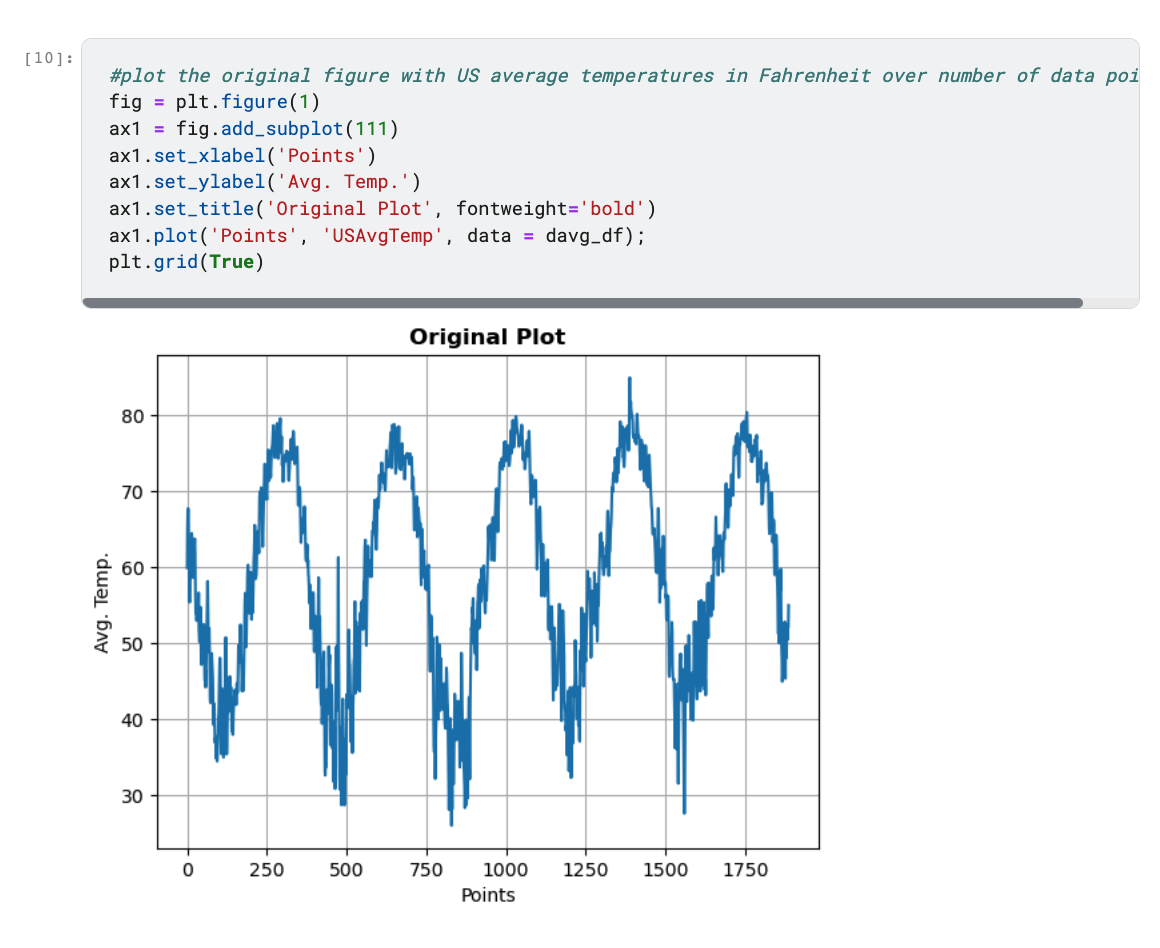
****

1. **Data Cleaning:**

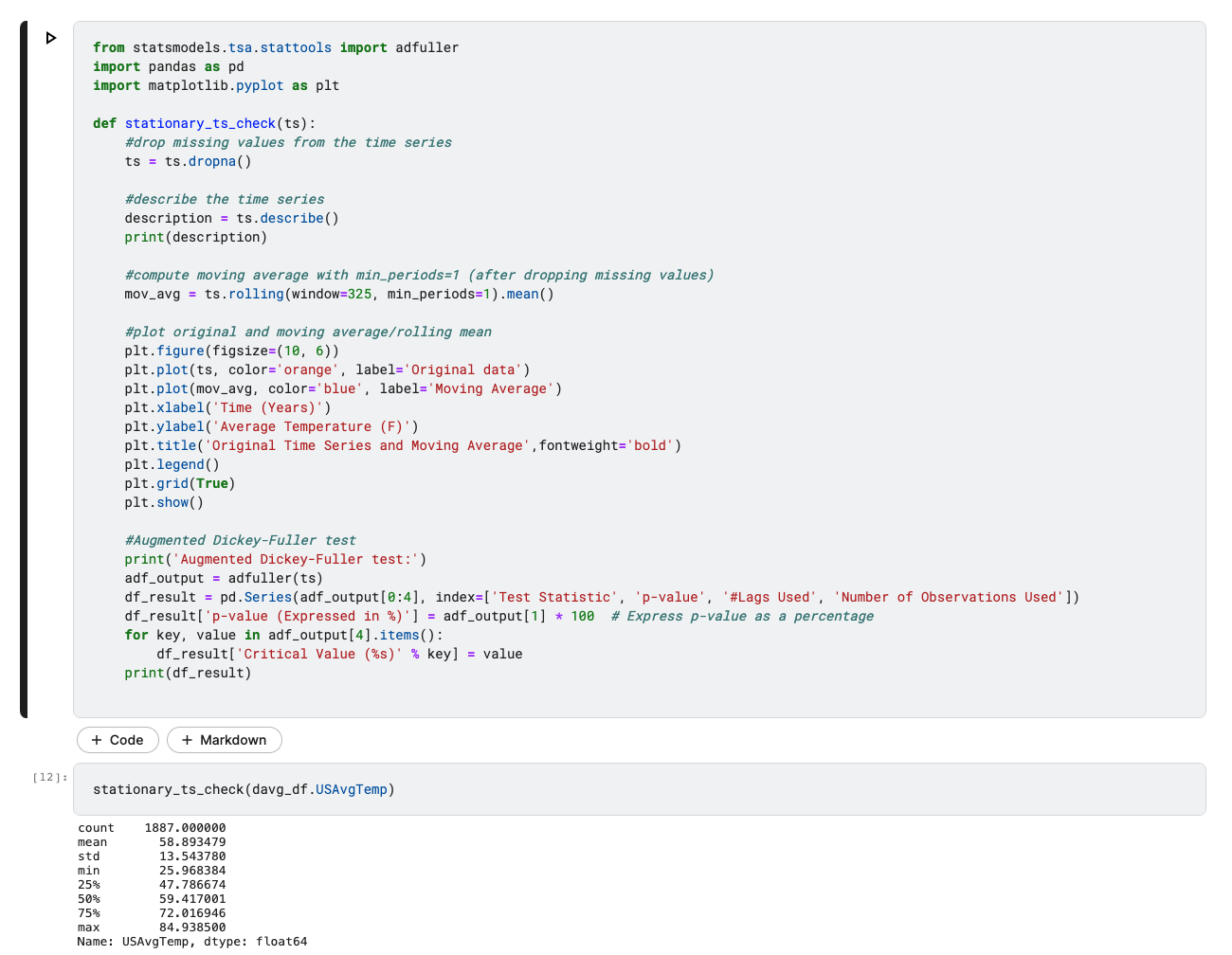
In addition to having excluded non-US cities, the data cleaning involved assigning points or ticks to better view the data without the context of time and also converting the measurements from kelvin to Fahrenheit for better interpretability.

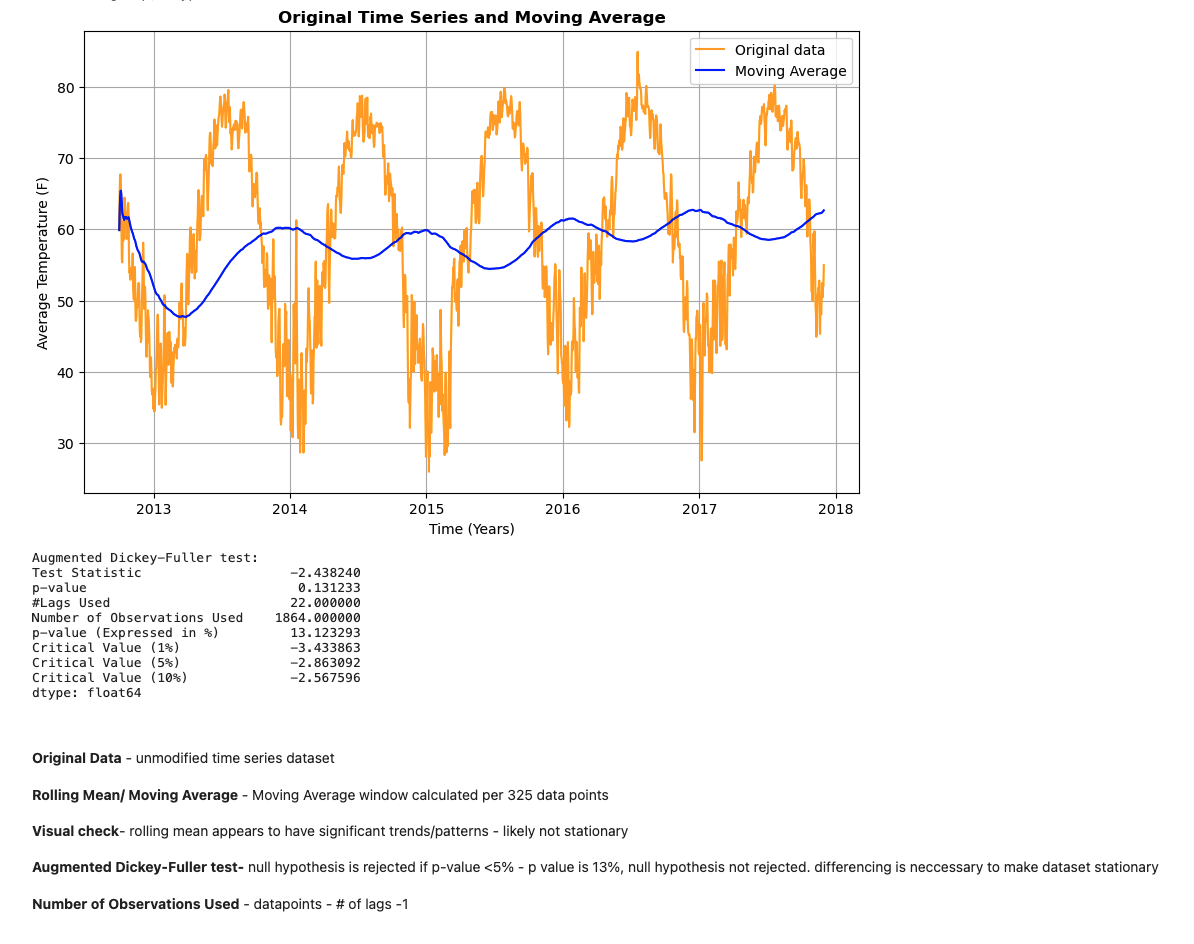
****

1. **Exploring and Visualization:**
2. Plot to see all unique data points

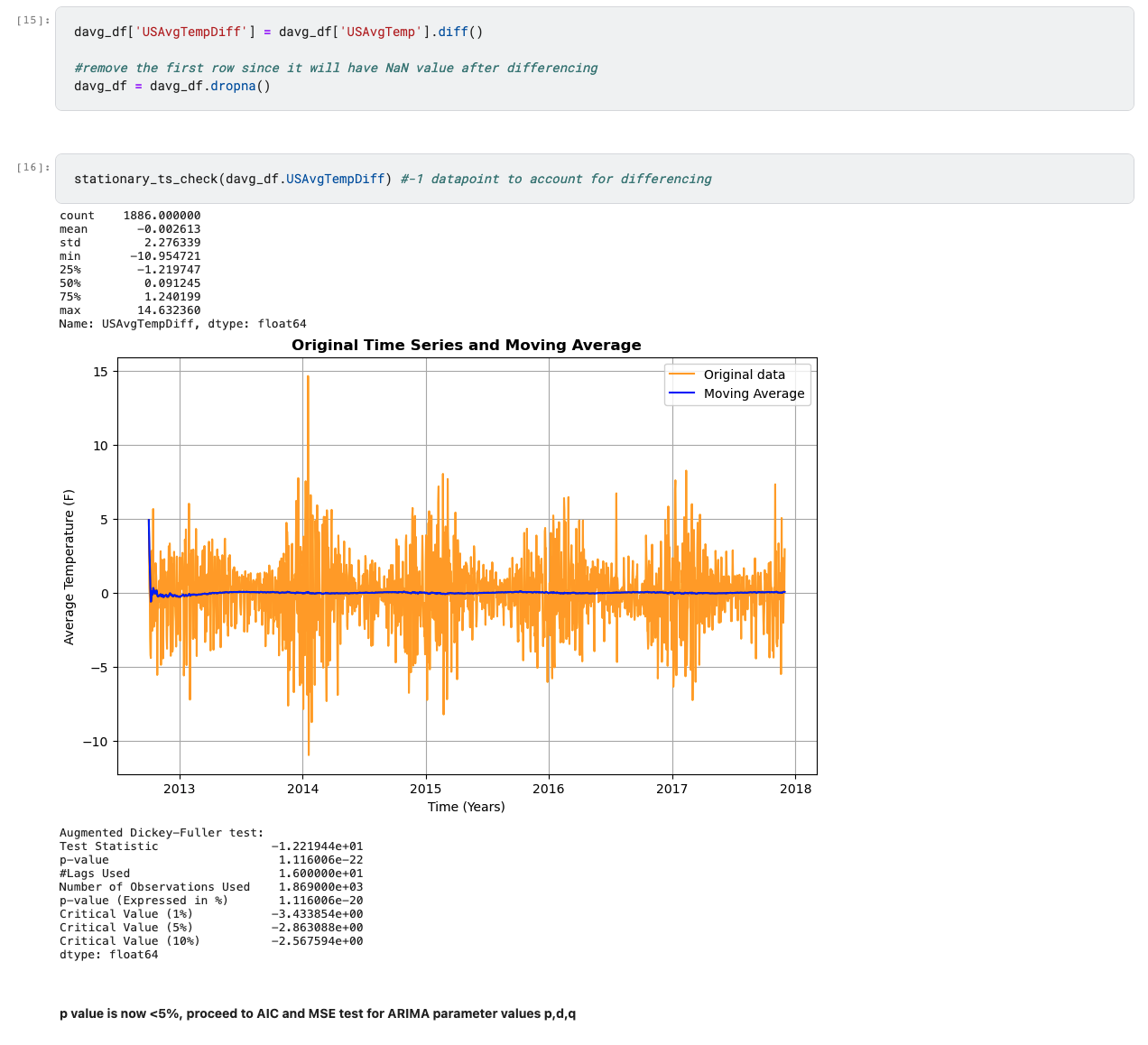


1. Perform Augmented Dickey-Fuller Test to determine if time series data is stationary. Reject null hypothesis if p-value less than 5%



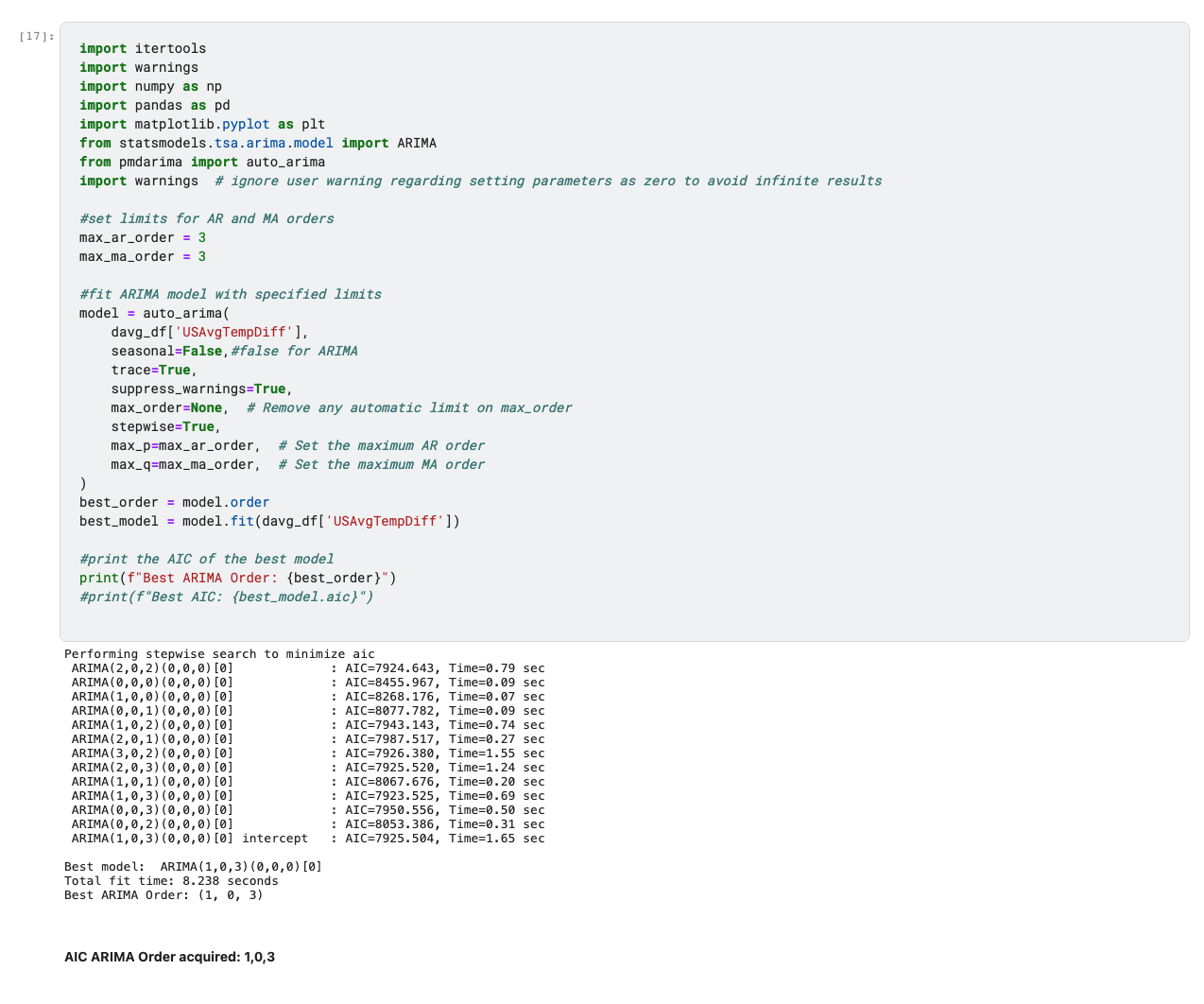


2a) If data is not stationary, difference and perform another ADF test until p value is less than 5%



**VI. Approach One: ARIMA Model**

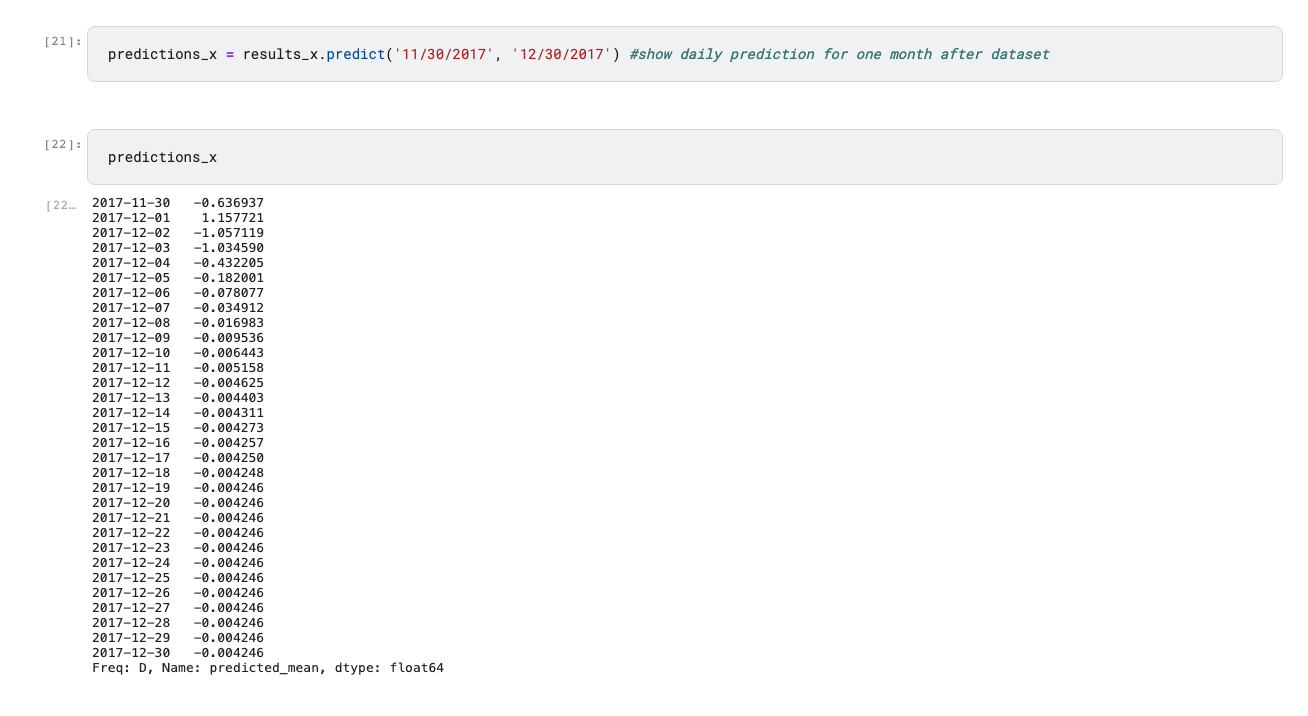
1. Perform AIC Test to acquire optimal ARIMA order



1. Perform MSE on ARIMA Fitted Values against original differenced dataset for validation

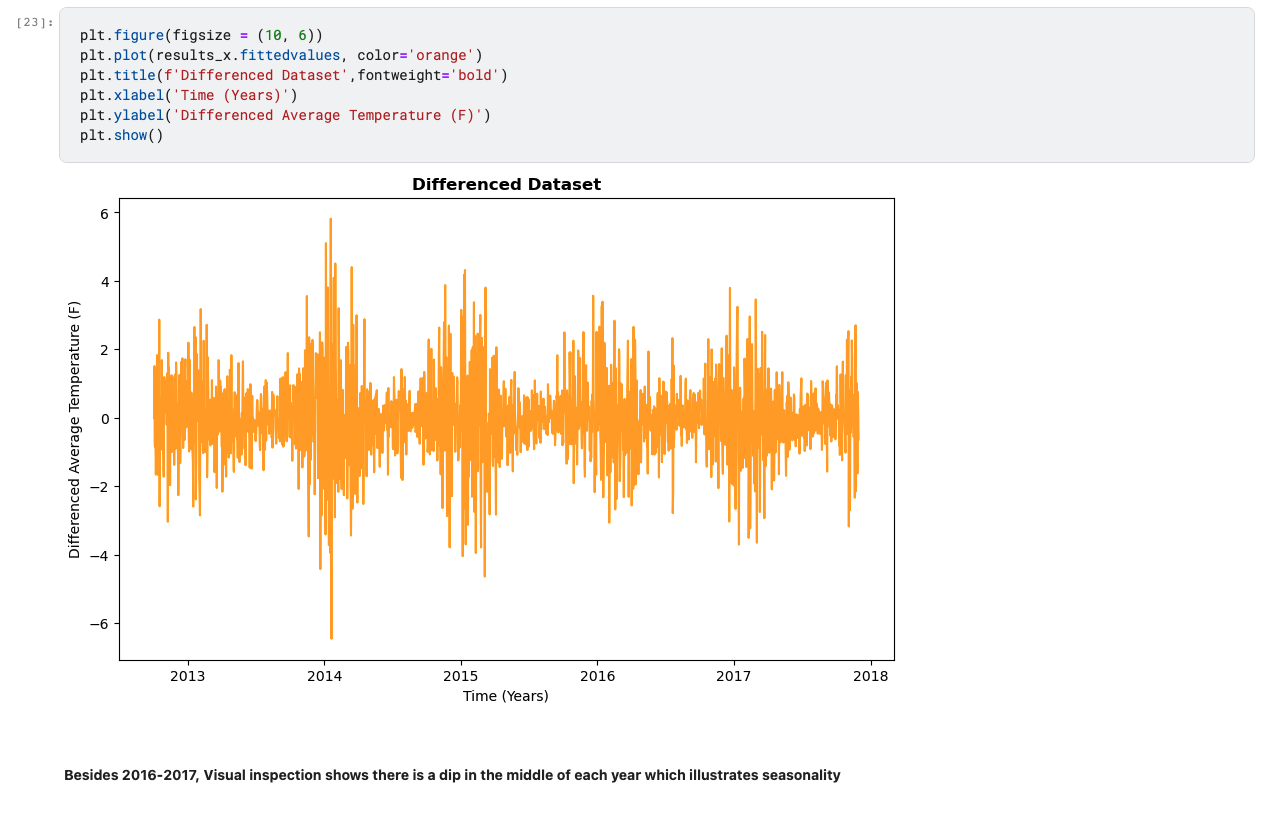


1. Show daily prediction for one month after dataset

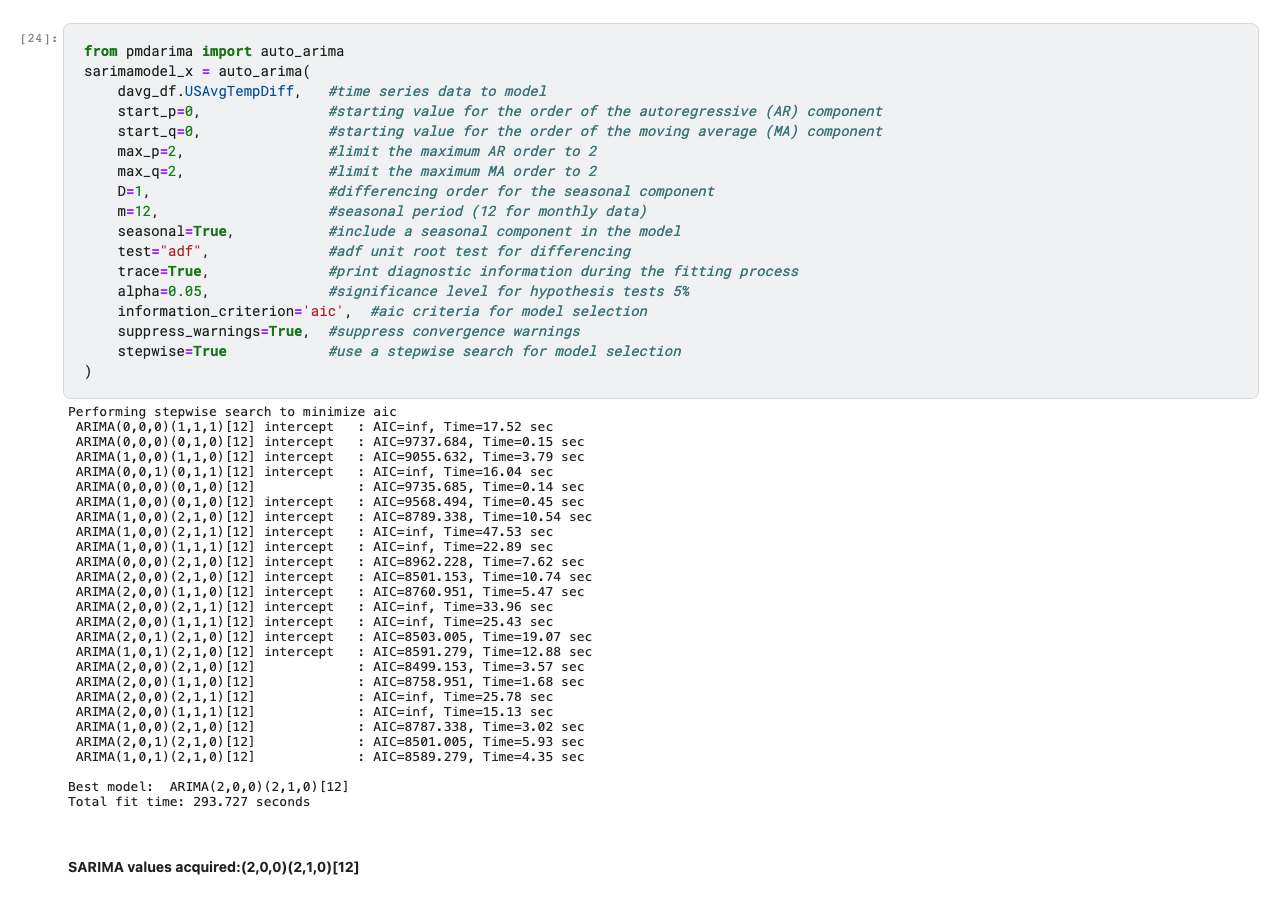


**VII. Approach Two: SARIMA Model**

1. Plot differenced dataset and check for seasonality by performing visual inspection on graph for seasonal pattern



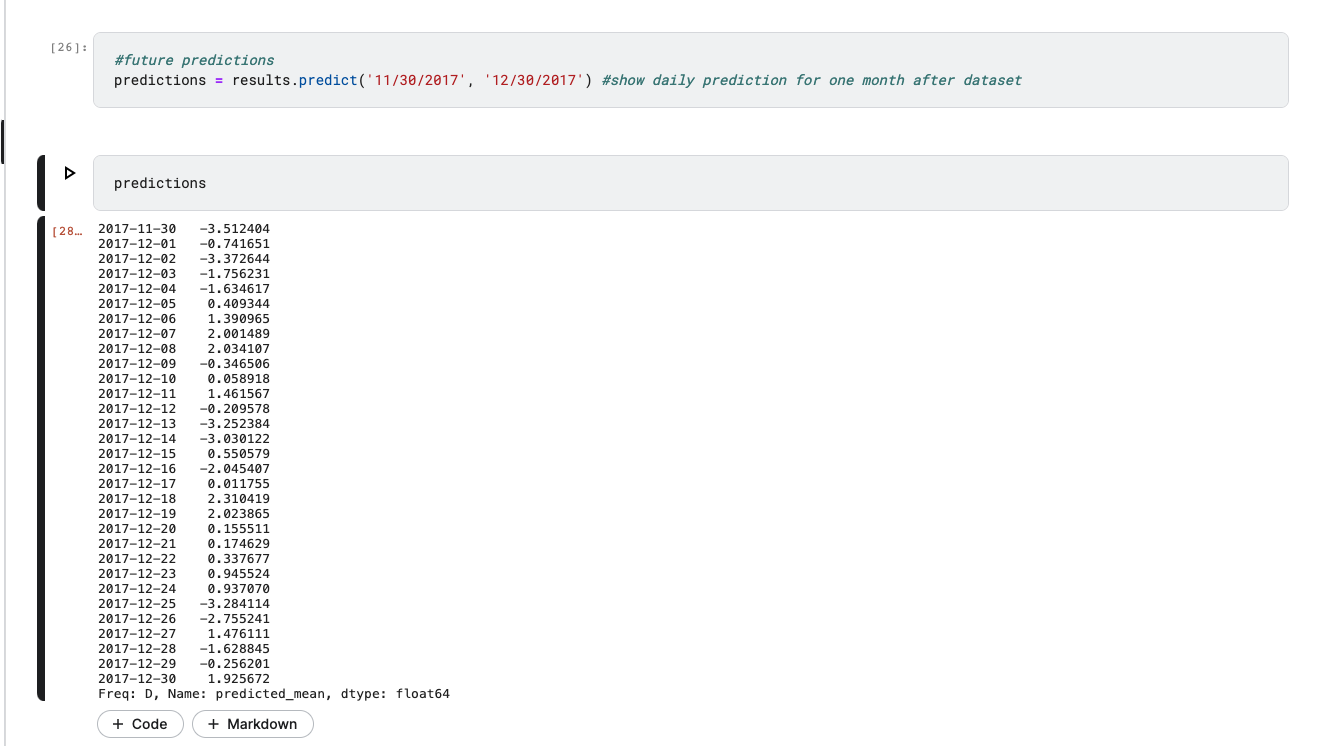
1. Perform AIC Test to acquire optimal SARIMA order



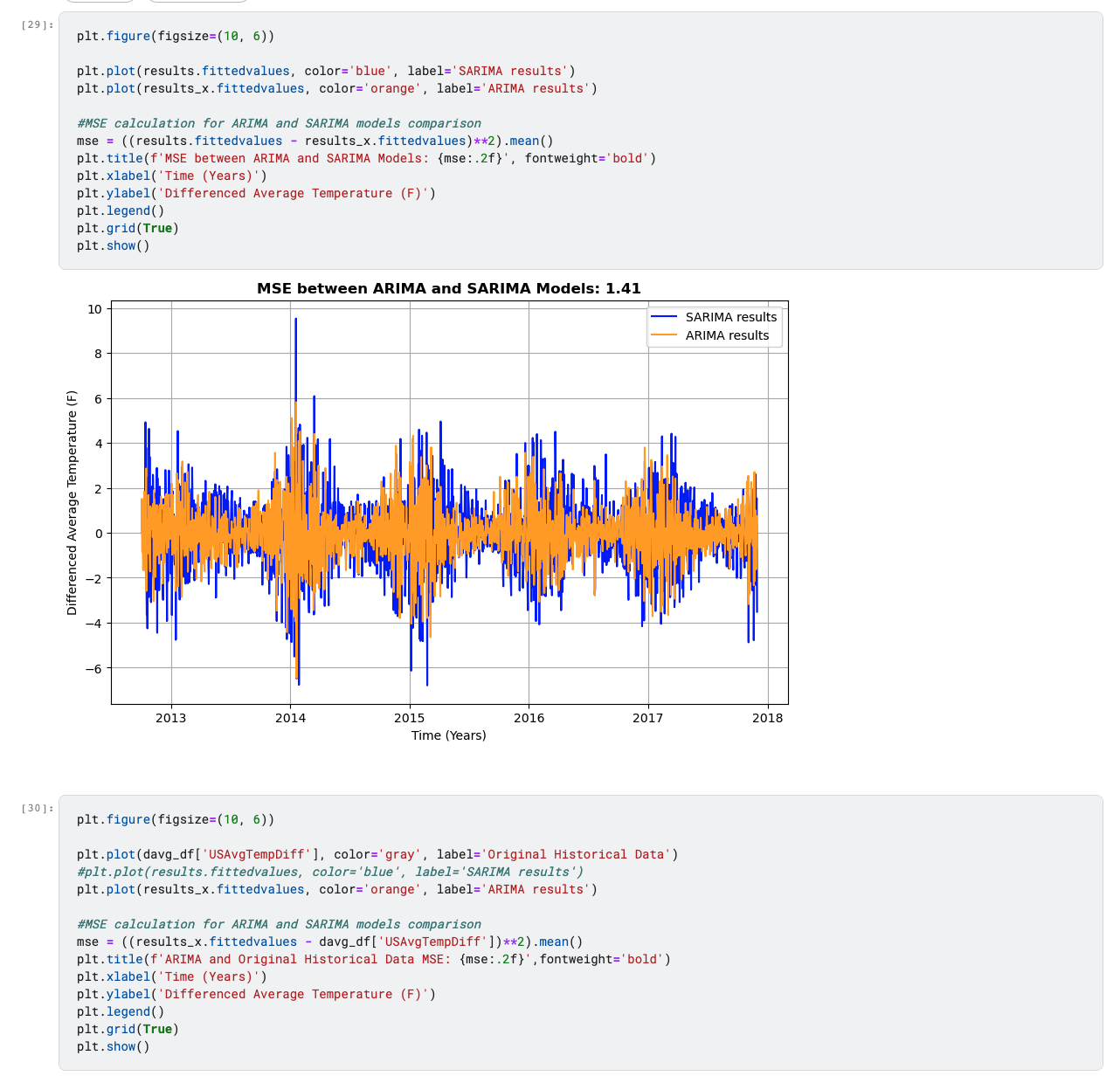
1. Perform MSE on SARIMA Fitted Values against original differenced dataset for validation

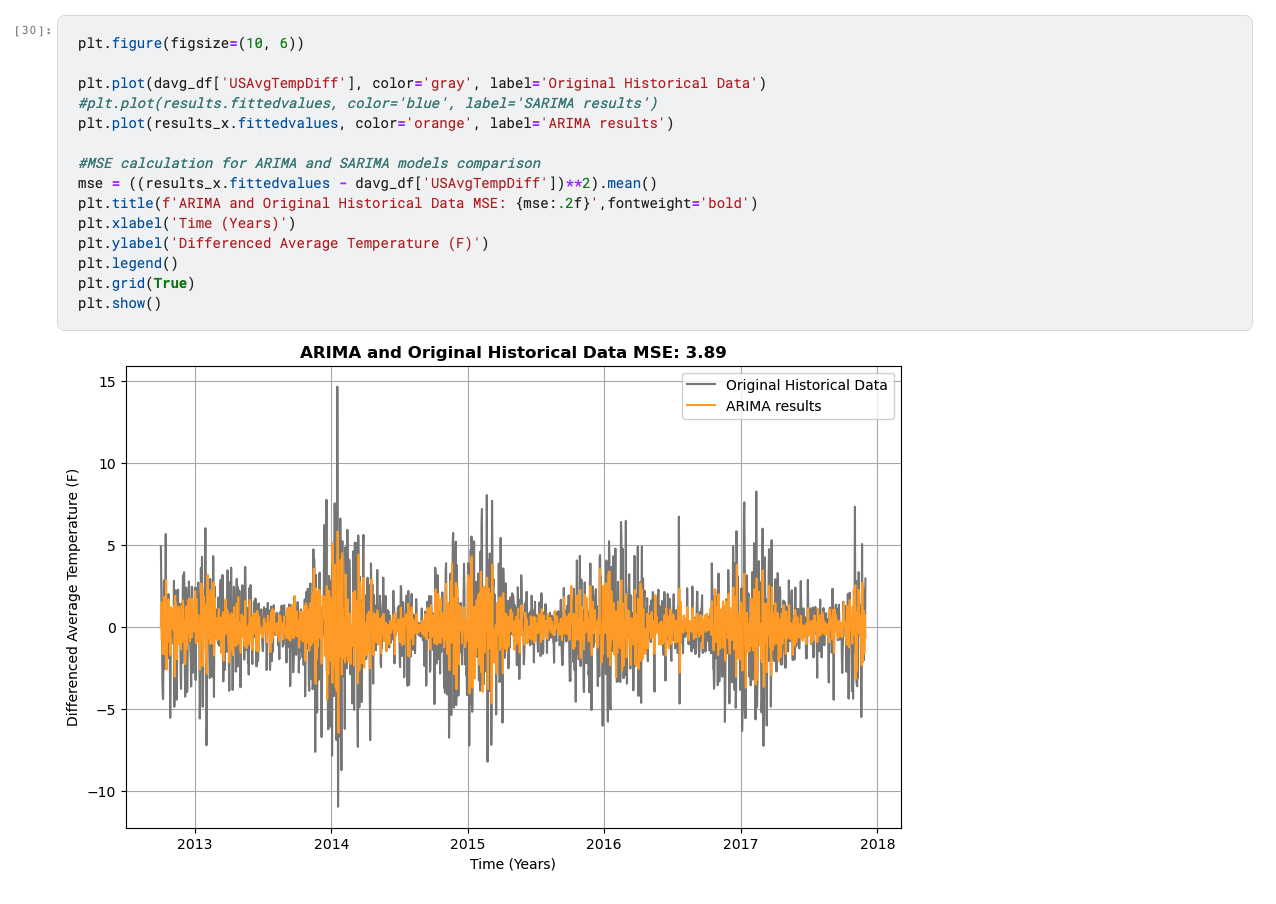


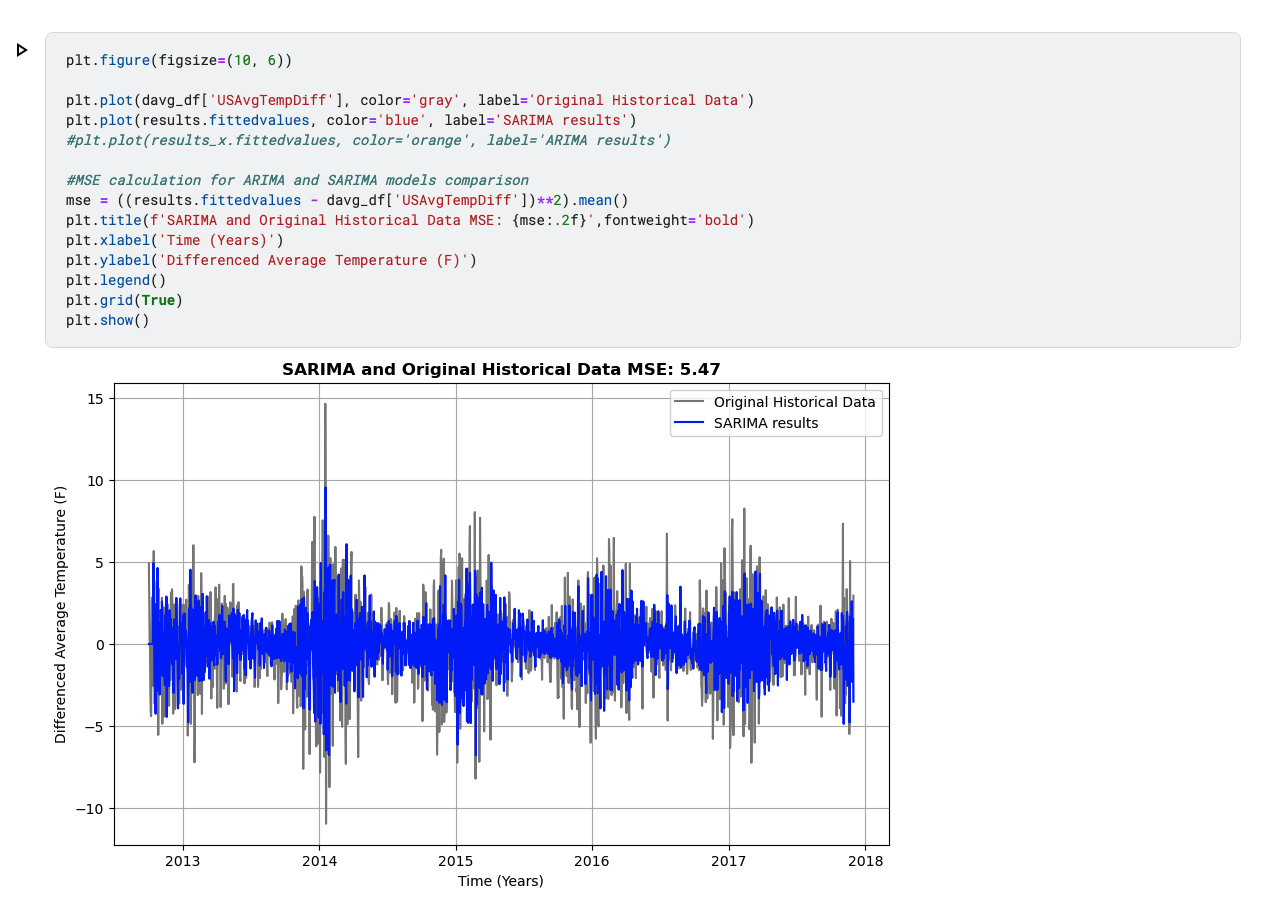
1. Show daily prediction for one month after dataset



**VIII. Validation for both Models via Mean Squared Error (MSE):**

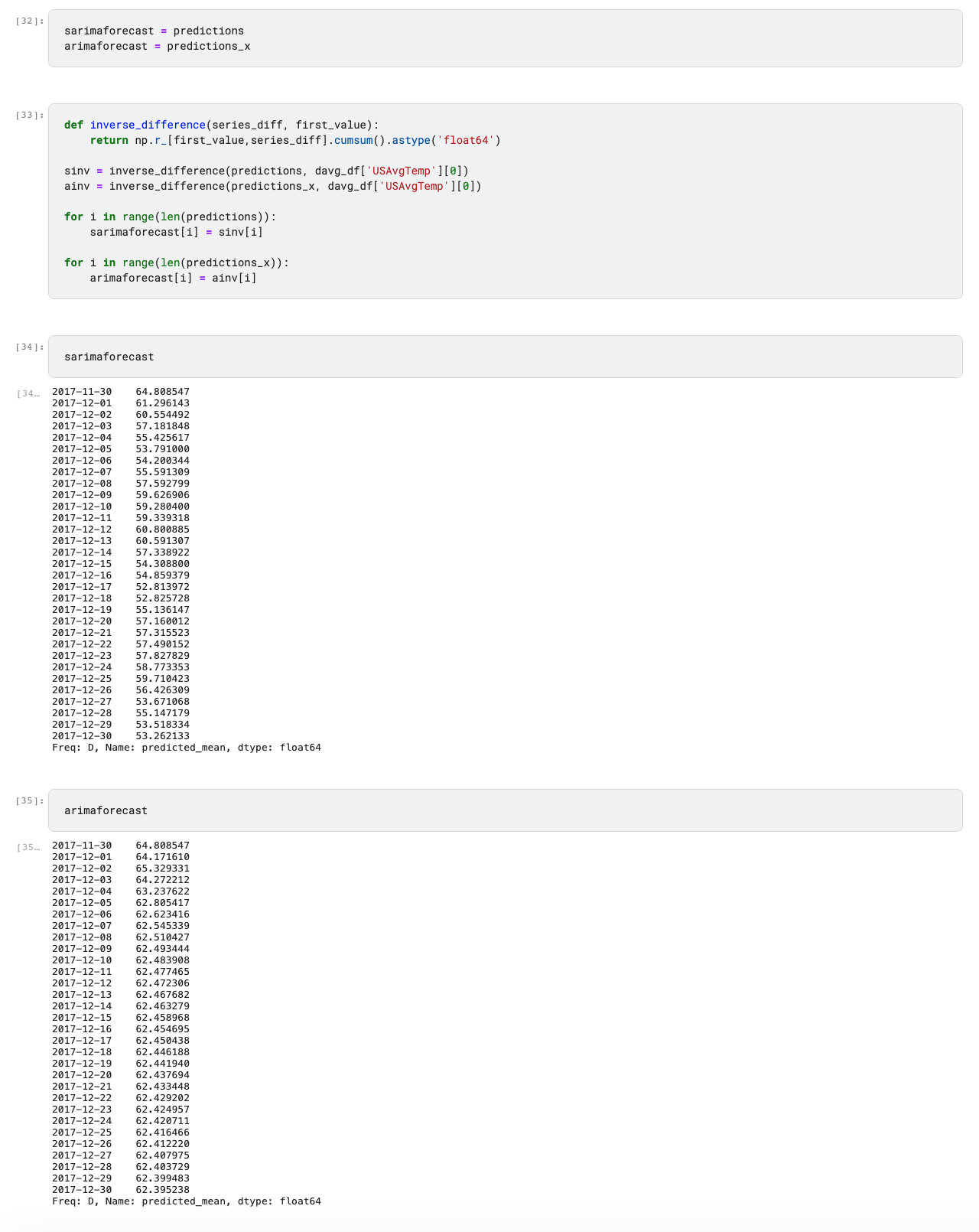
****

****

****

**IX. Final Results for both Models:**

1. Perform inverse operation on differenced dataset to revert data back to its original scale in Fahrenheit



1. Plot Historical Data and Predictions

